CAR-TR-833
CS-TR-3662
July 1996

N00014-96-1-0587
DAAH04-93-G-0419
IRI-9057934

Explaining Human Visual Space Distortion

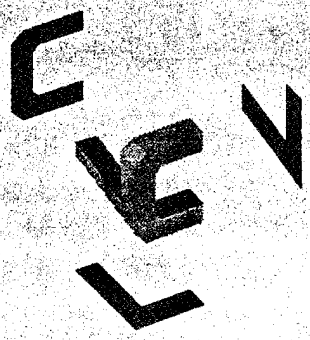
Cornelia Fermüller, LoongFah Cheong, and
Yiannis Aloimonos

Computer Vision Laboratory
Center for Automation Research
University of Maryland
College Park, MD 20742-3275

Abstract

: surrounded by surfaces that we perceive by visual means. Understanding the ba

COMPUTER VISION LABORATORY



CENTER FOR AUTOMATION RESEARCH

UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND
20742-3275

CAR-TR-833
CS-TR-3662
July 1996

N00014-96-1-0587
DAAH04-93-G-0419
IRI-9057934

Explaining Human Visual Space Distortion

Cornelia Fermüller, LoongFah Cheong, and
Yiannis Aloimonos

Computer Vision Laboratory
Center for Automation Research
University of Maryland
College Park, MD 20742-3275

Abstract

We are surrounded by surfaces that we perceive by visual means. Understanding the basic principles behind this perceptual process is a central theme in visual psychology, psychophysics and computational vision. Metric descriptions of physical space encoding distances between features in the environment have been used throughout the ages for various purposes. Naturally, such descriptions were used by early theorists for modelling perceptual space; that is, surfaces may be represented in our brains by encoding the distance of each point on the surface from our eye. The development of technology has allowed empirical scientists to perform accurate experiments measuring properties of perceptual space. It turns out that humans estimate a distorted version of their extra-personal space. A large number of experiments have been performed to study stereoscopic depth perception using tasks that involve the judgment of depth at different distances [8, 9, 13, 22]. Recently, a few experiments have been conducted to compare aspects of depth judgment due to stereoscopic and monocular motion perception [24]. In these experiments, it has been shown that from stereo vision humans over-estimate depth (relative to fronto-parallel size) at near fixations and under-estimate it at far fixations, whereas human depth estimates from visual motion are not affected by the fixation point. On the other hand, the orientation of an object in space does not affect depth judgment in stereo vision while it has a strong effect in motion vision, for the class of motions tested. This paper develops a computational geometric model that explains why such distortion might take place. The basic idea is that, both in stereo and motion, we perceive the world from multiple views. Given the rigid transformation between the views and the properties of the image correspondence, the depth of the scene can be obtained. Even a slight error in the rigid transformation parameters causes distortion of the computed depth of the scene. The unified framework introduced here describes this distortion in computational terms, in order to explain a number of recent psychophysical experiments on the perception of depth from motion or stereo.

The support of the Office of Naval Research under Grant N00014-96-1-0587, the National Science Foundation under Grant IRI-9057934, and the Army Research Office under Grant DAAH04-93-G-0419 is gratefully acknowledged. The second author was also supported in part by the Tan Kah Khee Post-graduate Scholarships.

1 Introduction

The nature of the representation of the world inside our heads as acquired by visual perception has persisted as a topic of investigation for thousands of years, from the works of Aristotle to the present [19]. In our day, answers to this question have several practical consequences in the field of robotics and automation. An artificial system equipped with visual sensors needs to develop representations of its environment in order to interact successfully with it. At the same time, understanding the way space is represented in the brains of biological systems is key to unravelling the mysteries of perception. We refer later to space represented inside a biological or artificial system as *perceptual space*, as opposed to *physical*, extra-personal *space*.

Interesting non-computational theories of perceptual space have appeared over the years in the fields of philosophy and cognitive science [17]. Computational theories, on the other hand, developed during the past thirty years in the area of computer vision, have followed a brute-force approach, equating physical space with perceptual space. Euclidean geometry involving metric properties has been used very successfully in modelling physical space. Thus, early attempts at modelling perceptual space concentrated on developing metric three-dimensional descriptions of space, as if it were the same as physical space. In other words, perceptual space was modelled by encoding the exact distances of features in three dimensions. The apparent ease with which humans perform a plethora of vision-guided tasks creates the impression that humans, at least, compute representations of space that have a high degree of generality; thus, the conventional wisdom that these descriptions are of a Euclidean metric nature was born and has persisted until now [1, 12, 19].

Computational considerations, however, can convince us that for a monocular or a binocular system moving in the world it is not possible to estimate an accurate description of three-dimensional metric structure, i.e., the exact distances of points in the environment from the nodal point of the eye or camera. This paper explains this in computational terms for the case of perceiving the world from multiple views. This includes the cases of both motion and stereo. Given two views of the world, whether these are the left and right views of a stereo system or successive views acquired by a moving system, the depth of the scene in view depends on two factors: (a) the three-dimensional rigid transformation between the views, hereafter called the *3D transformation*, and (b) the identification of image features in the two views that correspond to the same feature in the 3D world, hereafter called *visual correspondence*.

If there were no errors in the 3D transformation or the visual correspondence, then clearly the depth of the scene in view could be accurately recovered and thus a metric description could be obtained for perceptual space. Unfortunately, this is never the case. In the case of stereo, the 3D transformation amounting to the extrinsic calibration parameters of the stereo rig cannot be accurately estimated, only approximated [4]. In the case of motion, the three-dimensional motion parameters describing rotation and translation are estimated within error bounds [3, 5, 20, 26]. Finally, visual correspondence itself cannot be obtained perfectly; errors are always present. Thus, because of errors in both visual correspondence and 3D transformation, the recovered depth of the scene is always a *distorted* version of the scene structure. The fundamental contribution of

this paper is the development of a computational framework showing the geometric laws under which the recovered scene shape is distorted. In other words, there is a systematic way in which visual space is distorted; the transformation from physical to perceptual space belongs to the family of Cremona transformations [23].[1]

The power of the computational framework we introduce is demonstrated by using it to explain recent results in psychophysics. A number of recent psychophysical experiments have shown that humans make incorrect judgments of depth using either stereo [9, 13] or motion [24]. Our computational theory explains these psychophysical results and demonstrates that perceived space is not describable using a well-established geometry such as hyperbolic, elliptic, affine or projective. Understanding the invariances of distorted perceived space will contribute to the understanding of robust representations of shape and space, with many consequences for the problem of recognition. This work was motivated by our recent work on direct perception and qualitative shape representation [6, 7] and was inspired by the work of Koenderink and van Doorn on pictorial relief [16].

The organization of this paper is as follows. Section 2.1 introduces the concept of iso-distortion surfaces. Considering two close views, arising from a system in general rigid motion, we relate image motion measurements to the parameters of the 3D rigid motion and the depth of the scene. Then, assuming that there is an error in the rigid motion parameters, we find the computed depth as a function of the actual depth and the parameters of the system. Considering the points in space that are distorted by the same amount, we find them to lie on surfaces that in general are hyperboloids. These are the iso-distortion surfaces that form the core of our approach. In Section 2.2 we further describe the iso-distortion surfaces in both 3D and visual space and we introduce the concept of the holistic or H-surfaces. These are surfaces that describe all iso-distortion surfaces distorted by the same amount, irrespective of the direction (n_x, n_y) in the image in which measurements of visual correspondence are made. The H-surfaces are important in our analysis of the case of motion since measurements of local image motion can be in any direction and not just along the horizontal direction which is dominant in the case of stereo. Section 3 describes psychophysical experiments from the recent literature using motion and stereo, and Section 4 explains their results using the iso-distortion framework. Section 4.1 describes in detail the coordinate systems and the underlying rigid transformations for the specific experiments. Sections 4.2 and 4.3 explain the experimental results for motion and stereo respectively using the framework introduced here. The experiments on both motion and stereo chosen here were cleverly designed by Tittle et al. [24] so that the underlying geometries of the motion and stereo configurations are qualitatively similar. Thus, they are of great comparative interest. The computational arguments presented here are based on two key ideas. First, the 2D image representation derived for stereo perception is of a different nature than the one derived for motion perception. Second, the only thing assumed about the scene is that it lies in front of the image plane, and thus all depth estimates have to be positive; therefore, the percep-

[1] In the projective plane, a transformation $(x, y, z) \rightarrow (x', y', z')$ with $\rho x' = \phi_1(x, y, z)$, $\rho y' = \phi_2(x, y, z)$, $\rho z' = \phi_3(x, y, z)$ where ϕ_1, ϕ_2, ϕ_3 are homogeneous polynomials and ρ any scalar, is called a rational transformation. A rational transformation whose inverse exists and is also rational is called a Cremona transformation.

tual system, when estimating 3D motion, minimizes the number of image points whose corresponding scene points have negative depth values due to errors in the estimate of the motion. Section 5 concludes the paper and discusses the relationship of this work to other attempts in the literature to capture the essence of perceptual space.

2 Distortion of Visual Space

2.1 Iso-distortion Surfaces

As an image formation model, we use the standard model of perspective projection on the plane, with the image plane at a distance f from the nodal point parallel to the XY plane, and the viewing direction along the positive Z axis as illustrated in Figure 1. We want a model that can be used both for motion and stereo. Thus, we consider a differential model of rigid motion. This model is valid for stereo, which constitutes a special constrained motion, when making the small baseline approximation that is used widely in the literature [16].

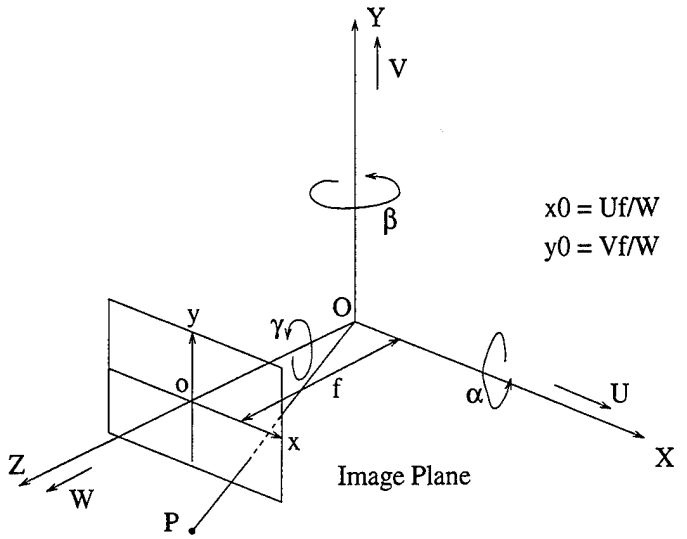


Figure 1: The image formation model. $OXYZ$ is a coordinate system fixed to the camera. O is the optical center and the positive Z-axis is the direction of view. The image plane is located at a focal length f pixels from O along the Z-axis. A point P at (X, Y, Z) in the world produces an image point p at (x, y) on the image plane where (x, y) is given by $\left(\frac{fX}{Z}, \frac{fY}{Z}\right)$. The instantaneous motion of the camera is given by the translational vector (U, V, W) and the rotational vector (α, β, γ).

The change of viewing geometry is described through a rigid motion with translational velocity (U, V, W) and rotational velocity (α, β, γ) of the observer in the coordinate system $OXYZ$.

As a consequence of the scaling ambiguity, only the direction of translation $(x_0, y_0) = \left(\frac{U}{W}f, \frac{V}{W}f\right)$ represented in the image plane by the epipole (also called the FOE (focus of expansion) or FOC (focus of contraction) depending on whether W is positive or

negative), the scaled depth Z/W and the rotational parameters can possibly be obtained from flow measurements. Using this notation the equations relating the 2D velocity $\mathbf{u} = (u, v) = (u_{\text{trans}} + u_{\text{rot}}, v_{\text{trans}} + v_{\text{rot}})$ of an image point to the 3D velocity and the depth of the corresponding scene point are

$$\begin{aligned} u &= u_{\text{trans}} + u_{\text{rot}} = (x - x_0)\frac{W}{Z} + \alpha xy - \beta\left(\frac{x^2}{f} + f\right) + \gamma y \\ v &= v_{\text{trans}} + v_{\text{rot}} = (y - y_0)\frac{W}{Z} + \alpha\left(\frac{y^2}{f} + f\right) - \frac{\beta xy}{f} - \gamma x \end{aligned} \tag{1}$$

where $u_{\text{trans}}, v_{\text{trans}}$ are the horizontal and vertical components of the flow due to translation, and $u_{\text{rot}}, v_{\text{rot}}$ the horizontal and vertical components of the flow due to rotation, respectively.

The velocity component $\mathbf{u}_n$ of the flow in any direction $\mathbf{n} = (n_x, n_y)$ has value

$$u_n = un_x + vn_y. \tag{2}$$

Knowing the parameters of the viewing geometry exactly, the scaled depth can be derived from (2). Since the depth can only be derived up to a scale factor, we set $W = 1$ and obtain

$$Z = \frac{(x - x_0)n_x + (y - y_0)n_y}{u_n - u_{\text{rot}}n_x - v_{\text{rot}}n_y}$$

If there is an error in the estimation of the viewing geometry, this will in turn cause errors in the estimation of the scaled depth, and thus a distorted version of space will be computed. In order to capture the distortion of the estimated space, we describe it through surfaces in space which are distorted by the same multiplicative factor, the so-called iso-distortion surfaces. To distinguish between the various estimates, we use the hat sign " ˆ " to represent estimated quantities, the unmarked letters to denote the actual quantities, and the subscript "ϵ" to represent errors, where the estimates are related as follows:

$$\begin{aligned} (\hat{x_0}, \hat{y_0}) &= (x_0 - x_{0_\epsilon}, y_0 - y_{0_\epsilon}) \\ (\hat{\alpha}, \hat{\beta}, \hat{\gamma}) &= (\alpha - \alpha_\epsilon, \beta - \beta_\epsilon, \gamma - \gamma_\epsilon) \\ \hat{\mathbf{u}}_{\text{rot}} &= (\hat{u}_{\text{rot}}, \hat{v}_{\text{rot}}) = \mathbf{u}_{\text{rot}} - \mathbf{u}_{\text{rot}_\epsilon} = (u_{\text{rot}} - u_{\text{rot}_\epsilon}, v_{\text{rot}} - v_{\text{rot}_\epsilon}) \end{aligned}$$

If we also allow for a noise term N in the estimate $\hat{u}_n$ of the component flow u_n, we have $\hat{u}_n = u_n + N$. The estimated depth becomes

$$\begin{aligned} \hat{Z} &= \frac{(x - \hat{x}_0)n_x + (y - \hat{y}_0)n_y}{\hat{u}_n - (\hat{u}_{\text{rot}}n_x + \hat{v}_{\text{rot}}n_y)} \quad \text{or} \\ \hat{Z} &= Z \cdot \left(\frac{(x - \hat{x}_0)n_x + (y - \hat{y}_0)n_y}{(x - x_0)n_x + (y - y_0)n_y + Z(u_{\text{rot}_\epsilon}n_x + v_{\text{rot}_\epsilon}n_y) + NZ}\right) \end{aligned} \tag{3}$$

From (3) we can see that $\hat{Z}$ is obtained from Z through multiplication by a factor given by the term inside the brackets, which we denote by D and call the distortion factor. In

the forthcoming analysis we do not attempt to model the statistics of the noise and we will therefore ignore the noise term. Thus, the distortion factor takes the form

$$D = \frac{(x-\hat{x}_0)n_x + (y-\hat{y}_0)n_y}{\begin{array}{l}(x-x_0)n_x + (y-y_0)n_y \\ +Z\left[\left(\frac{\alpha_\epsilon xy}{f} - \beta_\epsilon\left(\frac{x^2}{f}+f\right) + \gamma_\epsilon y\right)n_x + \left(\alpha_\epsilon\left(\frac{y^2}{f}+f\right) - \beta_\epsilon\frac{xy}{f} - \gamma_\epsilon x\right)n_y\right]\end{array}} \quad (4)$$

or, in a more compact form

$$D = \frac{(x-\hat{x}_0)\,n_x + (y-\hat{y}_0)\,n_y}{(x - x_0 + Zu_{\mathrm{rot}_\epsilon})\,n_x + (y - y_0 + Zv_{\mathrm{rot}_\epsilon})\,n_y}$$

Equation (4) describes, for any fixed direction (n_x, n_y) and any fixed distortion factor D, a surface $f(x, y, Z) = 0$ in space, which we call an iso-distortion surface. For specific values of the parameters $x_0, y_0, \hat{x}_0, \hat{y}_0, \alpha_\epsilon, \beta_\epsilon, \gamma_\epsilon$ and (n_x, n_y), this iso-distortion surface has the obvious property that points lying on it are distorted in depth by the same multiplicative factor D. Also, from (3) it follows that the transformation from perceptual to physical space is a Cremona transformation.

It is important to realize that, on the basis of the preceding analysis, the distortion of depth also depends upon the direction (n_x, n_y) and is therefore different for different directions of flow in the image plane. This means simply that if one estimates depth from optical flow in the presence of errors, the results can be very different depending on whether the horizontal, vertical, or any other component is used; depending on the direction, any value between $-\infty$ and $+\infty$ can be obtained! It is therefore imperative that a good understanding of the distortion function be obtained, before visual correspondences are used to recover the depth or structure of the scene.

In order to derive the iso-distortion surfaces in 3D space we substitute $x = \frac{fX}{Z}$ and $y = \frac{fY}{Z}$ in (4), which gives the following equation:

$$\begin{array}{l} D\left(\left(\alpha_\epsilon XY - \beta_\epsilon\left(X^2+Z^2\right) + \gamma_\epsilon YZ\right)n_x + \left(\alpha_\epsilon\left(Y^2+Z^2\right) - \beta_\epsilon XY - \gamma_\epsilon XZ\right)n_y\right) \\ \quad - \left(X - \frac{\hat{x}_0 Z}{f} - D\left(X - \frac{x_0 Z}{f}\right)\right)n_x - \left(Y - \frac{\hat{y}_0 Z}{f} - D\left(Y - \frac{y_0 Z}{f}\right)\right)n_y = 0 \end{array}$$

describing the iso-distortion surfaces as quadratic surfaces—in the general case, as hyperboloids. One such surface is depicted in Figure 2. Throughout the paper we will need access to the iso-distortion surfaces from two points of view. On the one hand we want to compare surfaces corresponding to the same D, but different gradient directions; thus we are interested in the families of D iso-distortion surfaces (see Figure 3a). On the other hand we want to look at surfaces corresponding to the same gradient direction $\mathbf{n}$, but different D's, the families of $\mathbf{n}$ iso-distortion surfaces (see Figure 3b). We will also be interested in the intersections of the surfaces with planes parallel to the XZ, YZ, and XY planes. These intersections give rise to families of iso-distortion contours; for an example see Figure 4.

2.2 Visualization of Iso-distortion Surfaces

The iso-distortion surfaces presented in the previous section were developed for the general case, i.e., when the 3D transformation between the views is a general rigid motion.

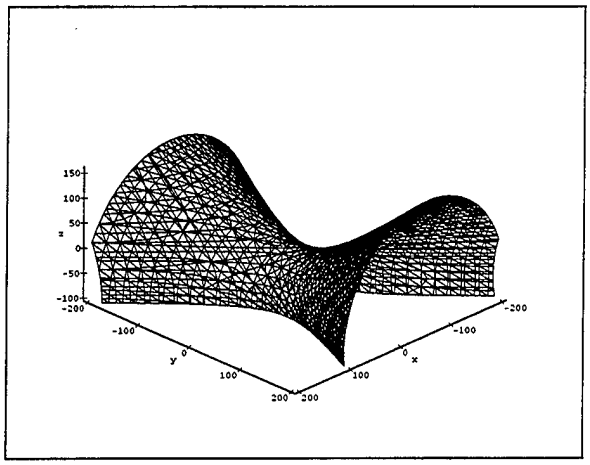


Figure 2: Iso-distortion surface in XYZ space. The parameters are: $x_0 = 10$, $x_{0_\epsilon} = -1$, $y_0 = -25$, $y_{0_\epsilon} = -5$, $\alpha_\epsilon = -0.05$, $\beta_\epsilon = -0.1$, $\gamma_\epsilon = -0.005$, $f = 1$, $D = 1.5$, $n_x = 0.7$.

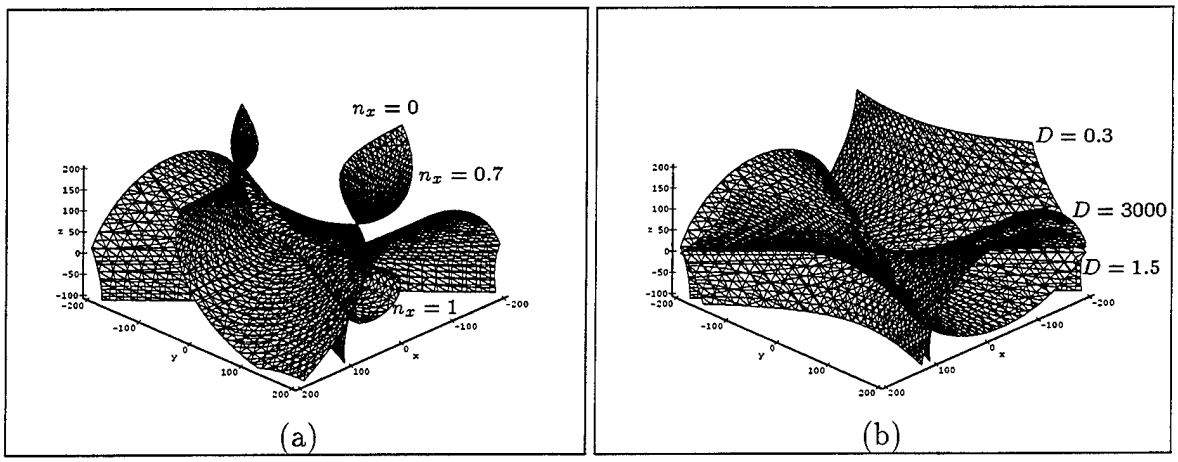


Figure 3: (a) Family of D iso-distortion surfaces for $n_x = 1, 0.7, 0$. (b) Family of $\mathbf{n}$ iso-distortion surfaces for $D = 0.3, 3000, 1.5$. The other parameters are as in Figure 2.

However, the psychophysical experiments that we will explain in the sequel considered constrained motion: rotation only around the Y-axis and translation only in the XZ plane. The only motion parameters to be considered are therefore β_ϵ, x_0 and $\hat{x}_0$, and the iso-distortion surfaces become

$$D\beta_\epsilon X^2 n_x + D\beta_\epsilon Z^2 n_x + D\beta_\epsilon XY n_y - (D-1)\, X n_x - (D-1)\, Y n_y - (\hat{x}_0 - D x_0)\, \frac{n_x}{f} Z = 0$$

which in general constitute hyperboloids. For horizontal flow vectors ($n_x = 1, n_y = 0$) they become elliptic cylinders and for vertical flow vectors they become hyperbolic cylinders.

Figure 5 provides an illustration of an iso-distortion surface for a general flow direction (here $n_x = 0.7$, $n_y = 0.714$). For our purposes, only the parts of the iso-distortion surfaces

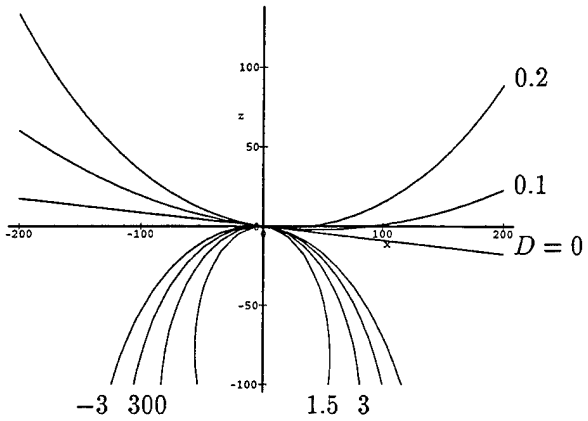
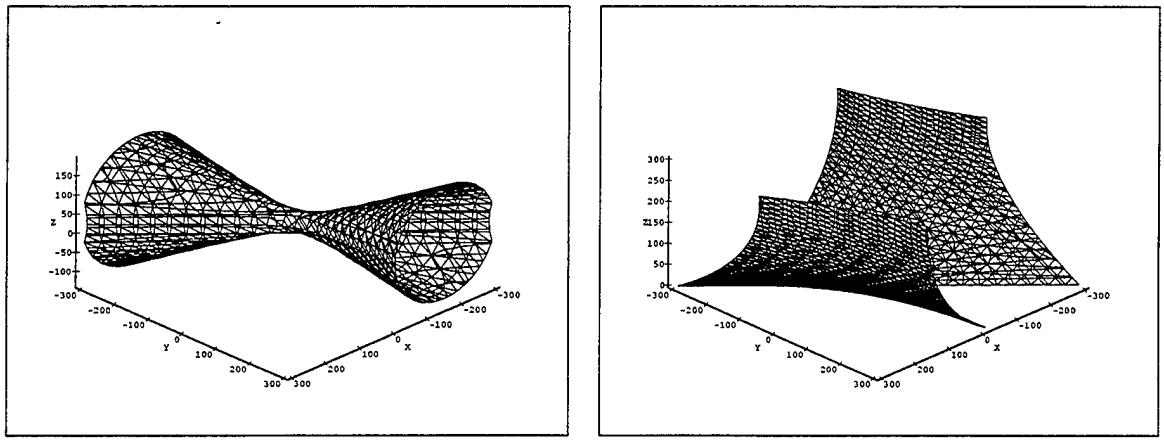


Figure 4: Intersection of a family of **n** iso-distortion surfaces (as shown in Figure 3b) with the XZ plane gives rise to a family of iso-distortion contours.

within the range visible from the observer are of interest. Since in the motion considered later the FOE has a large value, these parts show very little curvature and appear to be close to planar, as can be seen from Figure 5b.



(a) (b)

Figure 5: (a) A general iso-motion surface in 3D space. The Z-axis corresponds to the optical axis. (b) Section of an iso-motion surface for a limited field of view in front of the image plane for large values of x_0.

In order to make it easier to grasp the geometrical organization of the iso-distortion surfaces we next perform a simplification and use in addition to 3D space also visual space (that is, xyZ space): Within a limited field of view, terms quadratic in the image coordinates are small relative to linear and constant terms; thus we ignore them for the moment, which simplifies the rotational term for the motions considered to $(u_{\mathrm{rot}}, v_{\mathrm{rot}}) = (-\beta_\epsilon f, 0)$.

In visual space, i.e., xyZ space, that is the space perceived under perspective projection, where the fronto-parallel dimensions are measured according to their size on the

image plane, the iso-distortion surfaces take the following form:

$$[x(D-1) + (\hat{x}_0 - Dx_0)]\, n_x + y(D-1)n_y - D\beta_\epsilon f Z n_x = 0$$

That is, they become planes with surface normal vectors $((D-1)n_x, (D-1)n_y, -D\beta_\epsilon f n_x)$. For a fixed D, the family of D iso-distortion surfaces obtained by varying the direction (n_x, n_y) is a family of planes intersecting on a line l. If we slice these iso-distortion planes with a plane parallel to the xy (or image) plane, we obtain a pencil of lines with center lying on the x axis (the point through which line l passes) (see Figure 6a).

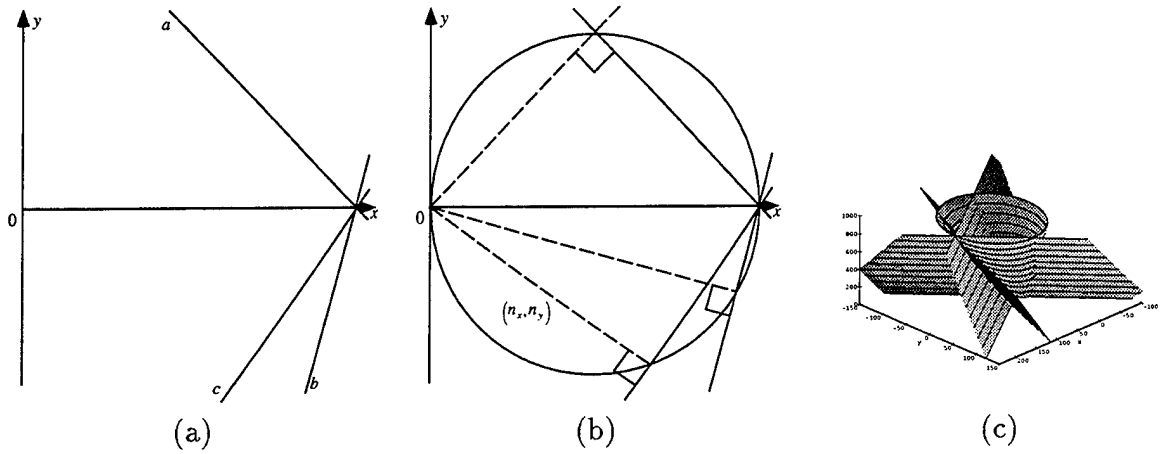


Figure 6: Simplified iso-distortion surfaces in visual space. (a) Intersection of the family of the simplified D iso-distortion surfaces (planes) for different directions (n_x, n_y) with a plane parallel to the image plane. (b) A circle represents the intersections of the family of the D iso-distortion surfaces with planes parallel to the image plane. (c) In visual space a family of D iso-distortion surfaces is characterized by a cone (the holistic surface).

In our forthcoming analysis we will need to consider the family of iso-distortion surfaces for a given distortion D, that is, the D iso-distortion surfaces for all directions (n_x, n_y). Thus, we will need a compact representation for the family of D iso-distortion surfaces in 3D space. The purpose of this representation is to visualize the high-dimensional family of D iso-distortion surfaces in $(x, y, Z, \mathbf{n})$ space through a surface in (x, y, Z) space in a way that captures the essential aspects of the parameters describing the family and thus the underlying distortion. As such a representation we choose the following surfaces, hereafter called holistic or H-surfaces, which are most easily understood through their cross sections parallel to the xy plane: Considering a planar slice of the family of D iso-distortion surfaces, as in Figure 6a, we obtain a pencil of lines. As a representation for these lines we choose the circle with diameter extending from the origin to the center of the pencil (Figure 6b). This circle clearly represents all orientations of the lines of the pencil (or the iso-distortion planes in the slicing plane). Any point P of the circle represents the slice of the iso-distortion plane which is perpendicular to a line through the center (O) and P.

If we now move the slicing plane parallel to itself, the straight lines of the pencil will trace the iso-distortion planes and the circle will change its radius and trace a circular cone with the Z axis as one ruling (Figure 6c).

The circular cones are described by the following equation:

$$x^2(D-1) + (\hat{x}_0 - Dx_0)\,x + y^2(D-1) - D\beta_\epsilon f Z x = 0$$

$$\text{or} \quad \left(x - \frac{(Dx_0 - \hat{x}_0 + D\beta_\epsilon f Z)}{2(D-1)}\right)^2 + y^2 = \left[\frac{D(x_0 + \beta_\epsilon f Z) - \hat{x}_0}{2(D-1)}\right]^2$$

Thus their axes are given by

$$Dx_0 - \hat{x}_0 + D\beta_\epsilon f Z - 2(D-1)x = 0, \quad y = 0$$

Slicing the cones and the simplified iso-distortion surfaces with planes parallel to the xy plane as in Figure 6b, the circles we obtain have center $(x, y, Z) = \left(\frac{Dx_0 - \hat{x}_0 + D\beta_\epsilon f Z}{2(D-1)}, 0, Z\right)$ and radius $\frac{D(x_0+\beta_\epsilon f Z)-\hat{x}_0}{2(D-1)}$. The circular cones serve as a holistic representation for the family of iso-distortion surfaces represented by the same D, therefore the name holistic or H-surface. It should be noted here that the holistic surfaces become cones only in the case of the constrained 3D motion considered in this paper. In the general case they are hyperboloids.

It must be stressed at this point that the iso-distortion surfaces should not be confused with the H-surfaces. Whereas a D iso-distortion surface for a direction $\mathbf{n}$ represents all points in space distorted by the same multiplicative factor D for image measurements in direction $\mathbf{n}$, the holistic surfaces do not represent any actually existing physical quantity; they serve merely as a tool for visualizing the family of D iso-distortion surfaces as $\mathbf{n}$ varies, and will be needed in explaining the distortion of space due to motion.

The H-surfaces for the families of iso-distortion surfaces vary continuously as we vary D. For $D = 0$ we obtain a cylinder with the Z-axis and the line $x = \hat{x}_0$ as diametrically opposite rulings. For $D = 1$ we obtain a plane parallel to the xy plane given by $Z = \frac{-x_{0\epsilon}}{\beta_\epsilon f}$; the cone for $D = \infty$ and the cone for $D = -\infty$ coincide. Thus we can divide the space into three areas: the areas between the $D = 0$ cylinder and the $D = -\infty$ cone, which only contain cones of negative distortion factor; the area between the $D = \infty$ cone and the $D = 1$ plane, with cones of decreasing distortion factor; and the area between the $D = 0$ cylinder and the $D = 1$ plane, with cones of increasing distortion factor. All the holistic surfaces intersect in the same circle, which is the intersection of the $D = 0$ cylinder and the $D = 1$ plane (see Figure 7a). Since the holistic surfaces intersect in one plane, any family of $\mathbf{n}$ iso-distortion surfaces intersects in a line in that plane.

To go back from visual to actual space, we have to compensate for the perspective scaling. In actual 3D space the iso-distortion surfaces are given by the equation

$$D\beta_\epsilon Z^2 n_x + (1-D)Xn_x + (1-D)Yn_x + (Dx_0 - \hat{x}_0)\frac{Zn_x}{f} = 0$$

describing parabolic cylinders curved in the Z dimension. Also the circular cones have an additional curvature in the Z dimension, and thus the H-surfaces in 3D space are surfaces of the form

$$X^2(D-1)f + Y^2(D-1)f + (\hat{x}_0 - Dx_0)\,XZ - D\beta_\epsilon X Z^2 f = 0$$

An illustration is given in Figure 7b.

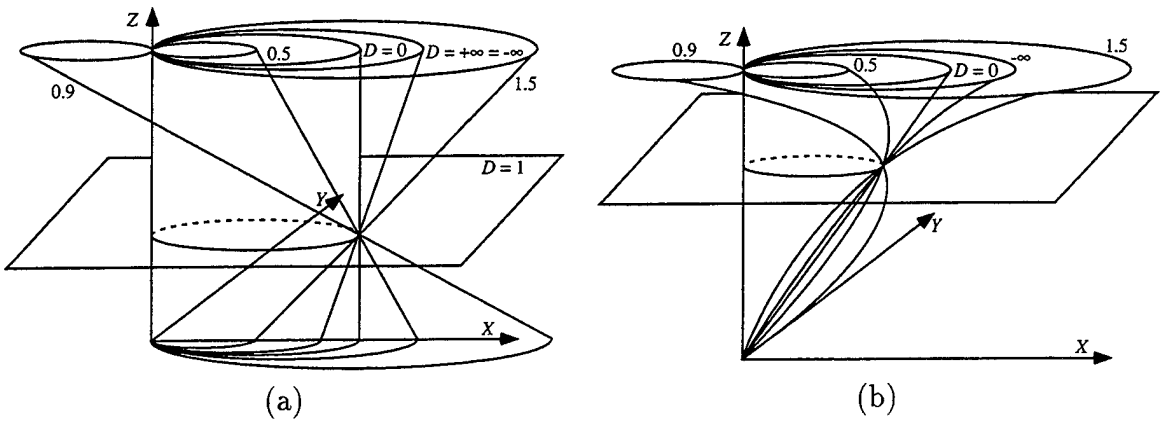


Figure 7: (a) Holistic surfaces (cones) in visual space, labeled with their respective distortion factors. (b) Holistic surfaces (third-order surfaces) in 3D space.

3 Psychophysical Experiments on Depth Perception

In the psychophysical literature a number of experiments has been reported that document a perception of depth which does not coincide with the actual situation. Most of the experiments were devoted to stereoscopic depth perception, using tasks that involved the judgment of depth at different distances. The conclusion usually obtained was that there is an expansion in the perception of depth of near distances and a contraction of depth at far distances. However, most of the studies did not explicitly measure perceived viewing distance, but asked for relative distance judgments instead. Recently a few experiments have been conducted by Tittle et al. [24] comparing aspects of depth judgment due to stereoscopic and monocular motion perception. The experiments were designed to test how the orientations of objects in space and their absolute distances influence the perceptual judgment. It was found that the stereoscopic cue and the motion cue give very different results.

The literature has presented a variety of explanations and proposed a number of models explaining different aspects of depth perception. Recently, great interest has arisen in attempts to explain the perception of visual space using well-defined geometries, such as similarity, conformal, affine, or projective transformations mapping physical space into perceived space, and it has been debated whether perceptual space is Euclidean, hyperbolic, or elliptic [27]. Our analysis shows that these models do not provide a general explanation for depth perception, and proposes that much of the data can be explained by the fact that the underlying 3D transformation is estimated incorrectly. Thus the transformation between physical and perceptual space is more complicated than previously thought. For the case of motion or stereo it is rational and belongs to the family of Cremona transformations [23].

We next describe a number of experiments and show that their results can be explained on the basis of imprecise estimation of the 3D transformation and thus can be predicted by the iso-distortion framework introduced here. Our primary focus in Sec-

tion 3.1 is on the experiments testing the difference between motion and stereo performed by Tittle et al. [24]. In addition, in Section 3.2 we describe two well-known stereoscopic experiments.

3.1 Distance Judgment from Motion and Binocular Stereopsis

In the first experiment [24] that we discuss, observers were required to adjust the eccentricity of a cylindrical surface until its cross-section in depth appeared to be circular. The observers could manipulate the cylindrical surface (shown in Figure 8) by rescaling it along its depth extent b (which was aligned with the Z-axis of the viewing geometry when the cylinder was in a fronto-parallel orientation) with the workstation mouse. Such a task requires judgment of relative distance. In order for the cross-section to appear circular, the vertical extent and the extent in depth of the cylinder, a and b, have to appear equal.

The experiments were performed for static binocular stereoscopic perception, for monocular motion, and for combined motion and stereopsis. The stereoscopic stimuli consisted of stereograms, and the monocular ones were created by images of cylinders rotating about a vertical axis (see Figure 8). In all the experiments the observers had to fixate on the front of the surface where it intersected the axis of rotation, and the cylindrical surfaces were composed of bright dots.

The effect of the slant and distance of the cylinder on the subjective depth judgment was tested. In particular, the cylinder had a slant in the range 0° to 30°, with 0° corresponding to a fronto-parallel cylinder as shown in Figure 8, and the distance ranged from 70 to 170 cm. Figure 9 displays the experimental results in the form of two graphs, with the x axis showing either the slant or distance and the y axis the adjusted eccentricity. An adjusted eccentricity of 1.0 corresponds to a veridical judgment, values less than this indicate an overestimate of b relative to a, and values greater than 1.0 indicate an underestimate. As can be seen from the graphs, whereas the perception of depth from motion only does not depend on the viewing distance, the extent b is overestimated for near distances and underestimated for far distances under stereoscopic perception. On the other hand, the slant of the surface has a significant influence on the perception of motion—at 0° b is overestimated and at 30° underestimated—and has hardly any influence on perception from stereo. The results obtained from the combined stereo and motion displays showed an overall pattern similar to those of the purely stereoscopic experiments.

For stereoscopic perception only, a very similar experiment, known as apparently circular cylinder (ACC) judgment, was performed in [9, 13], and the same pattern of results was reported there.

In a second experiment performed by Tittle et al. [24], the task was to adjust the angle between two connected planes until they appeared to be perpendicular to one another (see Figure 10).

Again the surfaces were covered with dots and the fixation point was at the intersection of the two planes and the rotation axis. As in the first experiment the influences of the cue (stereo, motion, or combined motion and stereo), the slant and the viewing distance on the depth judgment were evaluated. This task again requires a judgment of relative distance, that is, the depth extent b relative to the vertical extent a (as shown in

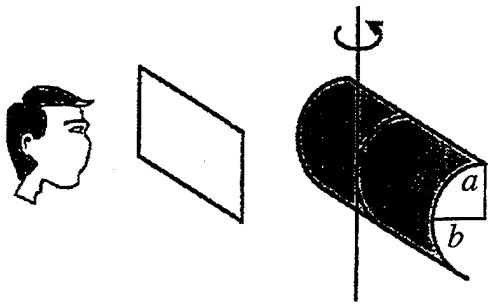


Figure 8: From [24]: a schematic view of the cylinder stimulus used in Experiment 1.

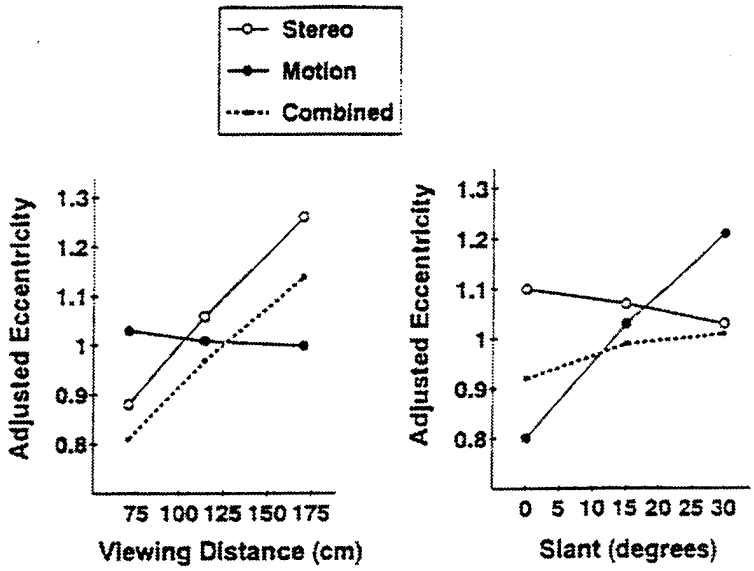


Figure 9: From [24]: Average adjusted cylinder eccentricity for the stereo, motion, and combined conditions as a function of simulated viewing distance and surface slant. An adjusted eccentricity of 1.0 indicates veridical performance.

Figure 10). The results displayed in Figure 11 are qualitatively similar to those obtained from the first experiment. An adjusted angle greater than the standard 90° corresponds to an overestimation of the extent in depth, and one less than 90° represents underestimation.

3.2 Stereoscopic Experiments: Apparent Fronto-parallel Plane/Apparent Distance Bisection

A classic test of depth perception for stereoscopic vision is the apparent fronto-parallel plane (AFPP) experiment [9, 22]. In this experiment, an observer views a horizontal array of targets. One target is fixed, usually in the median plane (Y-Z plane). The other targets are fixed in direction but are variable in radial distance under control of the subject. The subject sets these targets so that all of the targets appear to lie in

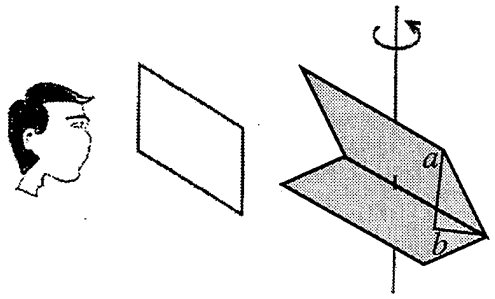


Figure 10: From [24]: a schematic view of the dihedral angle stimulus used in Experiment 2.

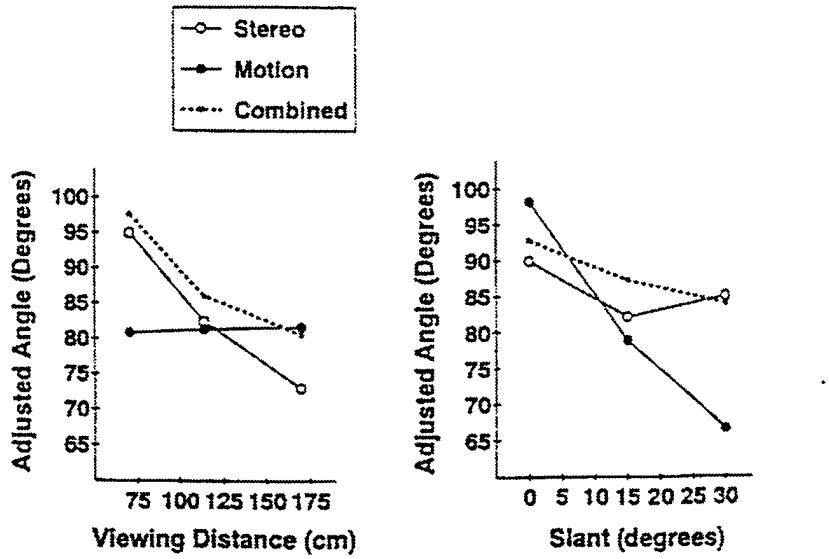


Figure 11: From [24]: Adjusted dihedral angle as a function of surface slant and simulated viewing distance. An adjusted angle of 90° indicates veridical performance.

a fronto-parallel plane. Care is taken so that fixation is maintained at one point. The results are illustrated in Figure 12.

The AFPP corresponds to a physical plane only at one distance, usually between 1m and 4m [9]. At far distances, the targets are set on a surface convex to the observer; at near distances they are set on a surface increasingly concave to the observer. Generally, the AFPP locus is skewed somewhat, that is, one side is farther away than the other.

In another classic experiment, instead of instructing a subject to set targets in an apparent fronto-parallel plane, the subjects are asked to set one target at half of the perceived distance of another target, placed in the same direction. This is known as the apparent distance bisection task or the ADB task [8]. In practice the targets would interfere with each other if they were in exactly the same direction, so they are displaced a few degrees. The task and the results are illustrated in Figure 13. These results were obtained with free eye movements, but the author claimed that the effect has also been

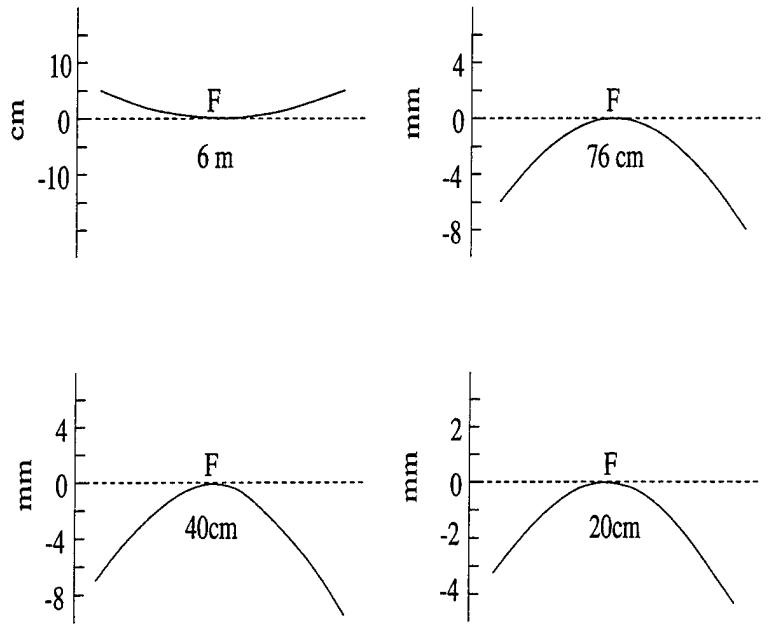


Figure 12: Data for the apparent fronto-parallel plane for different observation distances. In each case, F is the point of fixation. The visual field of the target extends from $-16°$ to $16°$. From [22].

replicated with fixation on one point.

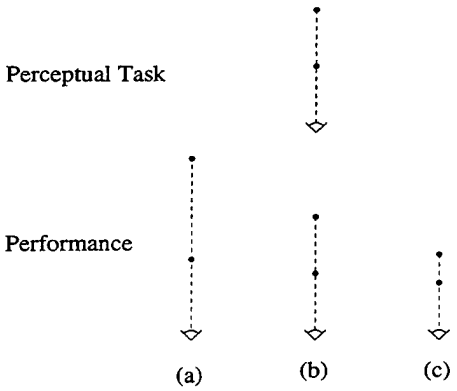


Figure 13: Apparent distance bisection task: (a) Far fixation point. (b) Correct distance judgment at intermediate fixation point. (c) Near fixation point.

4 Explanation of Psychophysical Results

4.1 The Viewing Geometry

(a) Stereo The geometry of binocular projection for an observer fixating on an environmental point is illustrated in Figure 14. We fix a coordinate system $(LXYZ)$ on the left eye with the Z-axis aligned with the optical axis and the Y-axis perpendicular to the fixation plane. In this system the transformation relating the right eye to the left eye is

a rotation around the Y-axis and a translation in the XZ plane. If we make the small baseline assumption, we can approximate the disparity measurements through a continuous flow field. The translational and rotational velocities are $(U, 0, W)$ and $(0, \beta, 0)$ respectively, and therefore the horizontal h and vertical v disparities are given by

$$\begin{aligned} h &= \frac{W}{Z}(x - x_0) \quad -\beta\left(\frac{x^2}{f} + f\right) \\ v &= \frac{W}{Z}y \qquad -\frac{\beta xy}{f} \end{aligned}$$

In the coordinate system thus defined (Figure 14), β is negative and x_0 is positive, and for a typical viewing situation very large. Therefore the epipole is far outside the image plane, which causes the disparity to be close to horizontal.

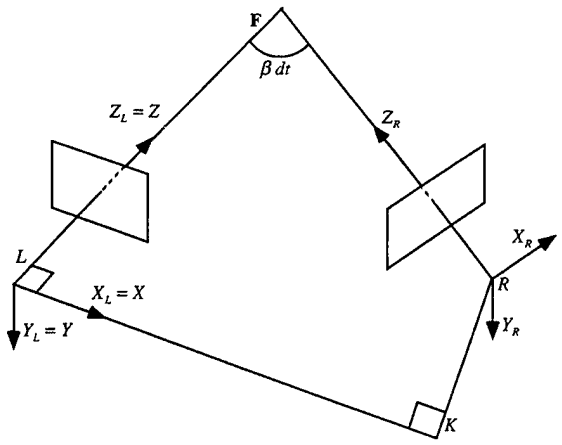


Figure 14: Binocular viewing geometry. $LK = U\,dt$ (translation along the X axis), $KR = W\,dt$ (translation along the Z axis), $LFR = \beta\,dt$ = convergence angle (resulting from rotation around the Y axis). L, K, R, F are in the fixation plane and dt is a hypothetical small time interval during which the motion bringing $X_L Y_L Z_L$ to $X_R Y_R Z_R$ takes place.

(b) Motion In the experiments described in Section 3.1 the motion of the object consists of a rotation around a vertical axis in space.

We fix a coordinate system to a point $S = (X_s, Y_s, Z_s)$ on the object in the YZ plane through which the rotation axis passes. At the time of observation it is parallel to the reference coordinate system $(OXYZ)$ on the eye of the observer (see Figure 15). In the new coordinate system on the object, the motion is purely rotational, and is given by the velocity $(0, w_y, 0)$. If we express this motion in the reference system as a motion of the observer we obtain a rotation around the Y-axis and an additional translation in the XZ-plane given by the velocity $(w_y Z_s, 0, -w_y X_s)$. Thus in the notation used before, there is a rotation with velocity $\beta = -w_y$, and a translation with epipole $(x_0, 0) = \left(-\frac{Z_s f}{X_s}, 0\right)$ or

$(\infty, 0)$ if $X_s = 0$. The value u_n of the flow component $\mathbf{u}_n$ along a direction $\mathbf{n} = (n_x, n_y)$ is given by

$$u_n = -w_y \left(\frac{X_s}{Z} \left(x + \frac{Z_s}{X_s} f \right) + \left(f + \frac{x^2}{f} \right) \right) n_x - w_y \left(\frac{y X_s}{Z} + \frac{xy}{f} \right) n_y$$

Since X_s is close to zero, x_0 again takes on very large values. In our coordinate system (see Figure 15) β is positive and x_0 is positive, since the circular cross-section is to the right of the YZ plane.

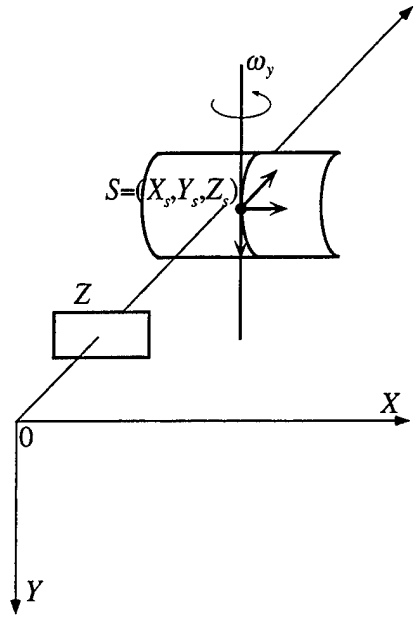


Figure 15:

Although the motion in the stereo and motion configurations is qualitatively similar, the psychophysical experimental results show that the system's perception of depth is not. This demonstrates that the two mechanisms of shape perception from motion and stereo work differently. We account for this by the fact that the 2D disparity representation used in stereo is of a different nature than the 2D velocity representation computed for further motion processing.

It is widely accepted that horizontal disparities are the primary input in stereoscopic depth perception although there have been many debates as to whether vertical disparities play a role in the understanding of shape [14, 21]. The fact is that for any human stereo configuration, even with fixation at nearby points, the horizontal disparities are much larger than the vertical ones. Thus, for the purpose of the forthcoming analysis, in the case of stereo we only consider horizontal disparities, although a small amount of vertical disparity would not influence the results.

On the other hand, for a general motion situation the actual 2D image displacements are in many directions. Due to computational considerations from local image measurements, only the component of flow perpendicular to edges can be computed reliably. This

is the so-called aperture problem. In order to derive the optical flow, further processing based on smoothing and optimization procedures has to be performed, which implicitly requires some assumptions about the smoothness of the scene. For this reason we expect the 2D image velocity measurements used by the system to be distributed in many directions, although the optical flow in the experimental motion is mostly horizontal.

Based on these assumptions about the velocity representations used, in the next two sections the experimental data—first the data from motion perception, then the data from stereo perception—are explained through the iso-distortion framework.

4.2 Motion

To visualize this and later explanations let us look at the possible distortions of space for the motion and stereo configurations considered here. Figure 16a gives a sketch of the holistic surfaces (third-order surfaces) for negative rotational errors (β_ϵ) and Figure 16b shows the surfaces for positive rotational errors. In both cases x_0 is positive. A change of the error in translation leaves the structure qualitatively the same; it only affects the sizes of the surfaces. In the overall pattern we observe a shift in the location of the intersection of the holistic surface. Since the intersection is in the $D = 1$ plane given by the equation $Z = -\frac{x_{0_\epsilon}}{\beta_\epsilon f}$, an increase in x_{0_ϵ} causes the intersection to have a larger Z coordinate in Figure 16a and a smaller one in Figure 16b. For both the motion and the stereo experiments, the FOE lies far outside the image plane. Therefore only a small part of the illustrated iso-distortion space actually lies in the observer's field of view. This part is centered around the Z-axis as schematically illustrated in Figure 16.

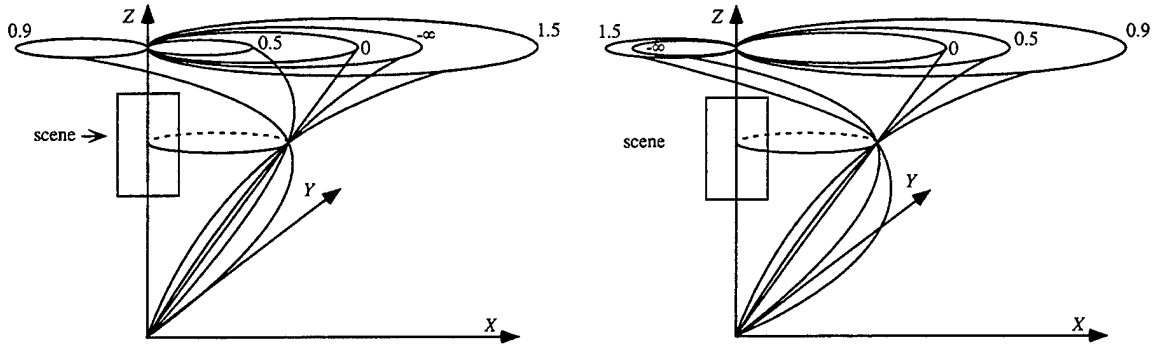


Figure 16: Holistic third-order surfaces for the geometric configurations described in the experiments. (a) Negative β_ϵ. (b) Positive β_ϵ.

The guiding principle in our explanation of the motion experiments lies in the minimization of negative depth estimates. We do not assume any scene interpretation; the only thing we know about the scene is that it lies in front of the image plane, and thus all depth estimates have to be positive. Therefore, we want to keep the number of image points, whose corresponding scene points would yield negative depth values due to erroneous estimation of the 3D transformation, as small as possible.

To represent the negative depth values we use a geometric statistical model: The scene in view lies within a certain range of depths between $Z_{\min}$ and $Z_{\max}$. The flow measurement vectors on the image are distributed in many directions; we assume that they follow some distribution. We are interested in the points in space for which we would estimate negative depth values.

For every direction $\mathbf{n}$ the points with negative depths lie between the $D = 0$ and $D = -\infty$ distortion surfaces within the range of depths covered by the scene. Thus, for every gradient direction we obtain a 3D subspace, covering a certain volume. The sum of all volumes for all gradient directions, normalized by the flow distribution considered here, represents a measure of the likelihood of negative depth estimates being derived from the image flow on the basis of some motion estimate. We call this sum the *negative depth volume.*

Let us assume there is some error in the estimate of the rotation, β_ϵ. We are interested in the translation error x_{0_ϵ} that will minimize the negative depth volume. Under the assumption that the distribution of flow directions is uniform (that is, the flow directions are uniformly distributed in every direction and at every depth within the range between $Z_{\min}$ and $Z_{\max}$), and that the simplified model is used (i.e., quadratic terms are ignored) and the computations are performed in visual space, the minimum occurs when the intersection of the iso-distortion cones is at the middle of the depth range of the scene. That is, the $D = 1$ plane is given as $Z = -\frac{x_{0_\epsilon}}{\beta_\epsilon f} = \frac{Z_{\min}+Z_{\max}}{2}$, and $x_{0_\epsilon} = -\beta_\epsilon f \frac{Z_{\min}+Z_{\max}}{2}$ [2].

Of course, we do not know the exact flow distribution, or the exact scene depth distribution, nor do we expect the system to optimally solve a minimization problem. We do, however, expect that the estimation of motion is such that the negative depth volume is kept rather small and thus that x_{0_ϵ} and β_ϵ are of opposite sign and the $D = 1$ plane is between the smallest and largest depth of the object observed.

In the following explanation we concentrate on the first experiment, which was concerned with the judgment about the circular cylinder.

We assume that the system underestimates the value of x_0, i.e., $x_{0_\epsilon} > 0$, because x_0 is very large and might even be infinite. Thus $\beta_\epsilon < 0$, and the distortion space of Figure 16b becomes applicable.

The holistic surfaces corresponding to negative iso-distortion surfaces in the field of view are very large in their circular extent, and thus the flow vectors leading to negative depth estimates are of large slope, close to the vertical direction. Figure 17 shows a cross-section through the negative iso-distortion surfaces and the negative holistic surfaces for a value Z in front of the $D = 1$ plane.

The rotating cylinder constitutes the visible scene. Its vertical cross-section along the axis of rotation lies in the space where x is positive. The most frontal points of the cross-section always lie in front of the $D = 1$ plane, and as the slant of the cylinder increases, the part of the cross-section which lies in front of the $D = 1$ plane increases as well.

The minimization of the negative depth volume and thus the estimation of the motion is independent of the absolute depth of the scene. Therefore a change in viewing distance should not have any effect on the depth perceived by the observer, *which explains the first experimental observation.*

The explanation of the second result lies in a comparison of the estimated vertical

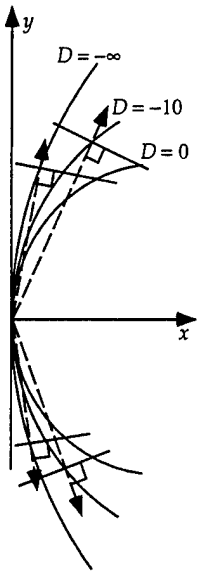


Figure 17: Cross-sections through negative iso-distortion surfaces and negative holistic surfaces. The flow vectors yielding negative depth values have large slopes.

extent, $\hat{a}$, and the extent in depth, $\hat{b}$.

Figures 18a–c illustrate the position of the circular cross-section in the distortion space for the fronto-parallel position of the cylinder. Section $a = (AC)$ lies at one depth and intersects the cross section of the holistic surface as shown in Figure 18b. Section $b = (BC)$ lies within a depth interval between depth values Z_B and Z_C. The cross-sections of the holistic surfaces are illustrated in Figure 18c. To make quantitative statements about the distortion D at any depth value, we assume that at any point P, D is the average value of all the iso-distortion surfaces passing through P. With this model we derive $\hat{a}$ and $\hat{b}$ as follows:

$$\hat{a} = Da \tag{5}$$

where D is the average distortion at the depth of section AC. The estimate $\hat{b}$ is derived as the difference of the depth estimate at points B and C. We denote by δ the difference between the average distortion factor of extent a and the distortion at point C, and we use ϵ to describe the change in the distortion factor from point C to point B. Thus

$$\begin{aligned} \hat{b} &= \hat{Z}_C - \hat{Z}_B \\ &= (D+\delta)Z_C - (D+\delta+\epsilon)(Z_C - b) \\ &= (D+\delta)b - \epsilon(Z_C - b) \end{aligned} \tag{6}$$

Z_C is much larger than b and thus $(Z_C - b)$ is always positive. Comparing equations (5) and (6) we see that for $a = b$ the factor determining the relative perceived length of a and b depends primarily on δ and ϵ.

For the case of a fronto-parallel cylinder, where extent a appears behind the $D = 1$ plane, δ is positive (see Figure 18b) and ϵ is negative (see Figure 18c), which means that b will be perceived to be greater than a.

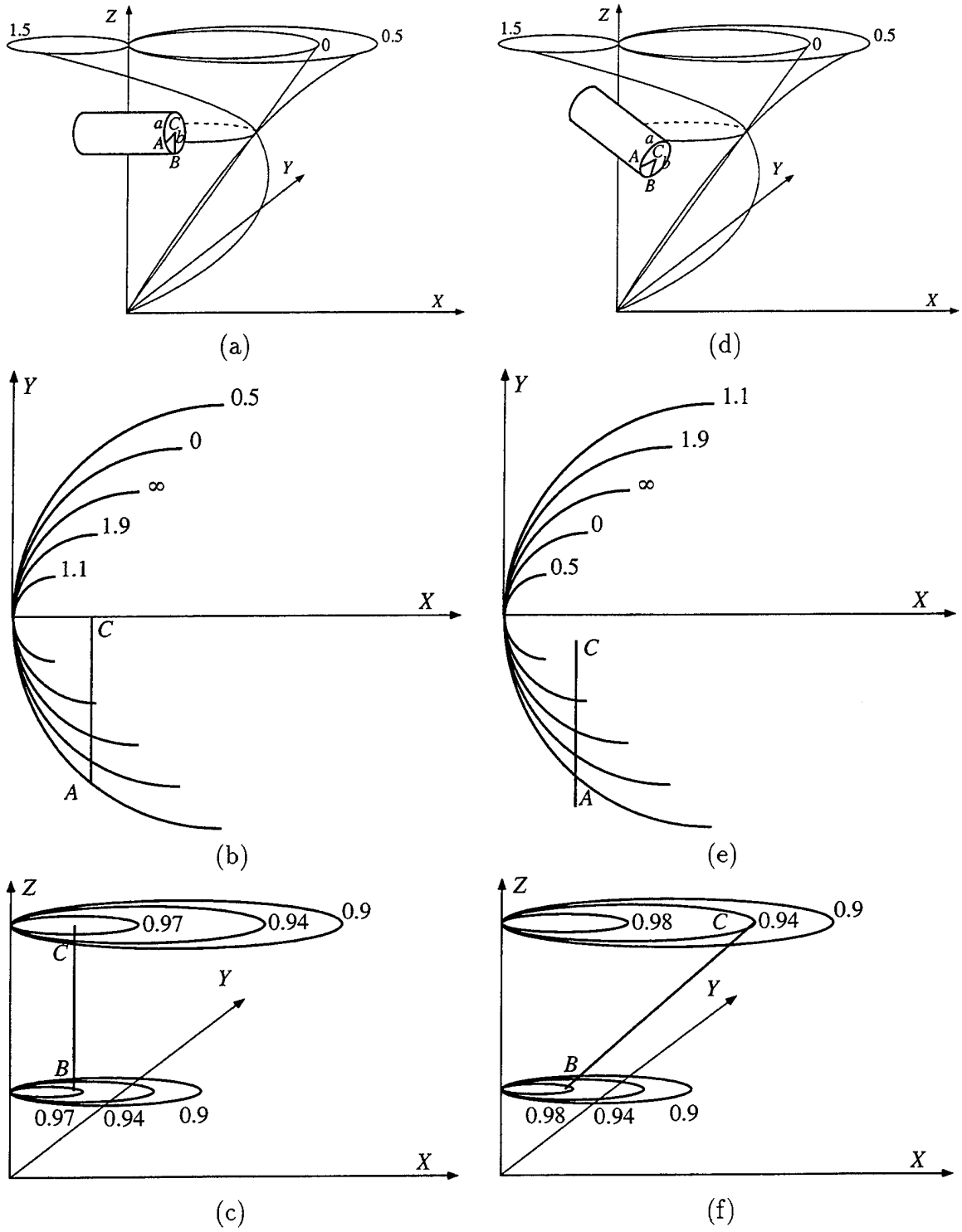


Figure 18: (a)–(c) Position of fronto-parallel cylinder in iso-distortion space. (d)–(f) Position of slanted cylinder in iso-distortion space. The figure shows that extent a appears behind the $D = 1$ plane in (b–c) and in front of the $D = 1$ plane in (e–f).

As the cylinder is slanted (see Figures 18d–f), the circular cross-section also becomes slanted. As a consequence the cylinder covers a larger depth range and extent a appears closer to or even in front of the $D = 1$ plane (see Figure 18e). Points on section b have increasing X-coordinates as Z increases (see Figure 18f). As the slant becomes large enough δ reaches a negative value, ϵ reaches a positive value and b is perceived to be smaller than a. Therefore the *results for the experiments involving the cylindrical surface* for the case of motion *can be explained* in terms of the iso-distortion diagrams with D that decreases or increases with Z.

The second experiment, concerned with the judgment of right angles, can be explained by the same principle. The estimate is again based on judgment of the vertical extent a relative to the extent in depth b (see Figure 10). *Either* we encounter the situation where the sign of x_0 is positive, so that a and b are measured mostly to the right of the YZ plane, and Figure 16b explains the iso-distortion space; *or* x_0 is negative, so that a and b are mostly to the left of the YZ plane, and the iso-distortion space is obtained by reflecting the space of Figure 16b in the YZ plane. In both cases the explanation given for the first experiment still applies. Due to the changes of position of the two planes in iso-distortion space with a change in slant, the extent in depth will be overestimated for the fronto-parallel position and underestimated for larger slants.

4.3 Stereo

In the case of stereoscopic perception the primary 2D image input is horizontal disparity. Due to the far-off location of the epipole the negative part of the distortion space for horizontal vectors does not lie within the field of view, as can be seen from Figure 16.

Since depth estimation in stereo vision has long been of concern to researchers in psychophysics, a large amount of experimental data has been published, and the parameters of the human viewing geometry are well documented. In [9] Foley studied the relationship between viewing distance and error in the estimation of convergence angle (β in our notation). From experimental data he obtained the relationship between perceived convergence angle and actual convergence angle shown in Figure 19.

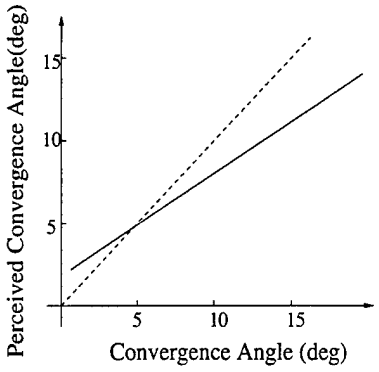


Figure 19: Perceived convergence angle as a function of convergence angle.

According to his data, the convergence angle is overestimated at far distances and underestimated at near distances. Foley expressed the data through the following rela-

tionship:

$$-\hat{\beta} = E + G(-\beta)$$

with E and G in the vicinity of 0.5; in the figures displayed here the following parameters based on data of Ogle [22] have been chosen: $E = 0.91°$ and $G = 0.66°$.

On the basis of these data, models have been proposed [8, 9, 22] that explain the perception of concavity and convexity for objects in a fronto-parallel plane. To account for the skewing described in the AFPP task the ocular images have been assumed to be of different sizes.

In our explanation based on the iso-distortion framework we make use of the experimental data of Figure 19 to explain β_ϵ. For far fixation points β_ϵ is negative and the iso-distortion space of Figure 16a applies. If we also take into account the quadratic term in the horizontal disparity formula of Section 4.1(a) (that is, the rotational part $\beta_\epsilon(\frac{x^2}{f} + f)$), we obtain an iso-distortion configuration for horizontal vectors as shown in Figure 20. In particular Figure 20a shows the contours obtained by intersecting the iso-distortion surfaces with planes parallel to the xZ plane in visual space, and Figure 20b shows the same contours in actual 3D space. Irrespective of x_{0_ϵ} the iso-distortion factor decreases with depth Z. The sign of x_{0_ϵ} determines whether the $D = 1$ contour (the intersection of the $D = 1$ surface with the xZ plane) is in front of or behind the image plane, and the exact position of the object with regard to the $D = 1$ contour determines whether the object's overall size is over- or underestimated.

For near fixation points, β_ϵ is positive and the iso-distortion space appears as in Figure 16b. The corresponding iso-distortion contours derived by including the quadratic term are illustrated in Figure 20c and d.

The perceived estimates $\hat{a}$ and $\hat{b}$ are modelled as before. However, this time it is not necessary to refer to an average distortion D, since only one flow direction is considered. Section a lies in the yZ plane and $\hat{a}$ is estimated as aD, with D the distortion factor at point C. The estimate for b is

$$\hat{b} = Db - \epsilon(Z_C - b)$$

As can be seen from Figures 20a and c, ϵ is increasing if the fixation point is distant and decreasing if the fixation point is close, and we thus obtain the under- and overestimation of $\hat{b}$ as experimentally observed. A slanting of the object has very little effect on the distortion pattern because the fixation point is not affected by it. As long as the slant is not too large, causing ϵ to change sign, the qualitative estimation of depth should not be affected by a change in slant. The slant might, however, influence the amount of over- and underestimation. There should be a decrease in the estimation error as the slant increases, since section b covers a smaller range of the distortion space. This can actually be observed from the experimental data in Figure 9.

The same explanation covers the second experiment related to the judgment of angles.

The iso-distortion patterns outlined here also explain the purely stereoscopic experiments. With regard to the AFPP task it can be readily verified that the iso-distortion diagram of Figure 20a (far fixation point) causes a fronto-parallel plane to appear on a concave surface, and thus influences the observer to set them at a convex AFPP locus, whereas the diagram of Figure 20c (near fixation point) influences the observer to

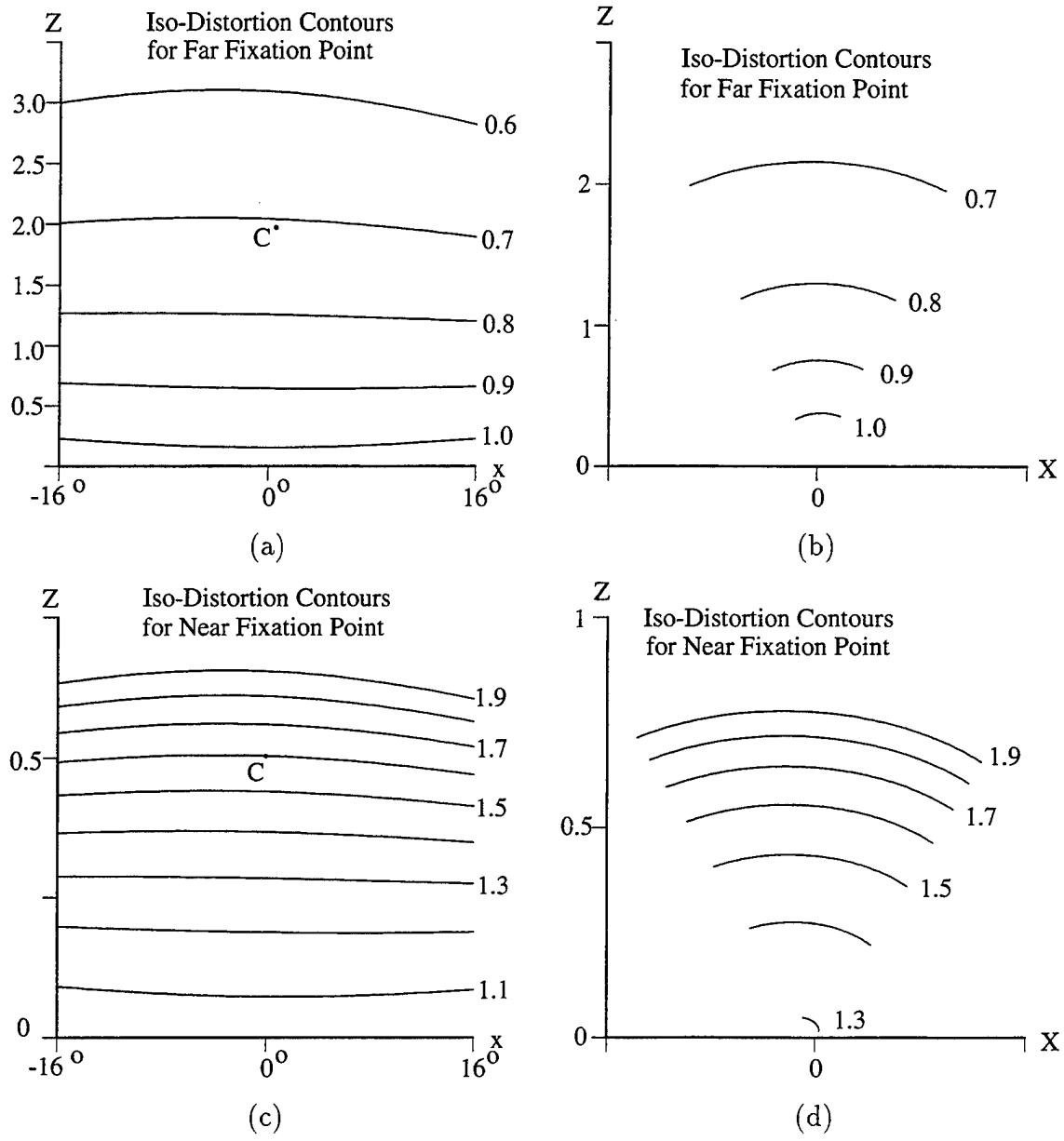


Figure 20: Iso-distortion contours for horizontal disparities: (a, b) for far fixation point in visual space (a) and actual space (b); (c, d) for near fixation point in visual and actual space.

set them on a concave AFPP locus. In addition, the skewing of the AFPP loci is also predicted by the iso-distortion framework.

Finally, with regard to the ADB task, the iso-distortion patterns predict that the target will be set at a distance closer than half-way to the fixation point if the latter is far, and at a distance further than half-way to the fixation point if the latter is near, which is in agreement with the results of the task.

5 Conclusions

The geometric structure of the visual space perceived by humans has been a subject of great interest in philosophy and perceptual psychology for a long time. With the advent of digital computers and the possibility of constructing anthropomorphic robotic devices that perceive the world in a way similar to the way humans and animals perceive it, computational studies are beginning to be devoted to this problem [15].

Many synthetic models have been proposed over the years in an attempt to account for the systematic distortion between physical and perceptual space. These range from Euclidean geometry [10] to hyperbolic [18] and affine [25] geometry. Many other interesting approaches have also been proposed, such as the Lie group theoretical studies of Hoffman [11] and the work of Koenderink and van Doorn [16], that are characterized by a deep geometric analysis attempting to discover invariant quantities of the distorted perceptual space under some assumed model. It is generally believed in the biological sciences that a large number of shape representations are computed in our heads and different cues are processed with different algorithms. For the case of motion and/or stereo, there might exist more than one process performing local analysis of motion or stereo disparity. The analysis proposed here has concentrated on a global examination of motion or disparity fields to explain a number of psychological results about the distortion of visual space that takes place over an extended field of view.

In contrast to the synthetic approaches in the literature, we have offered an analytic account of a number of properties of perceptual space. Our starting point was the fact that when we have multiple views of a scene (motion or stereo), then the 3D rigid transformation relating the views, and functions of local image correspondence, determine the perceived depth of the scene. However, even slight miscalculations of the parameters of the 3D transformation result in computing a distorted version of the actual physical space. In this paper, we studied geometric properties of the computed distorted space. The transformation between physical and perceptual space (i.e., actual and computed space) is a Cremona transformation. We have concentrated on analyzing the distortions from first principles, through an understanding of iso-distortion loci. The analytic geometric framework we have introduced is adequate for computationally explaining a set of psychophysical experiments related to the perception of shape from either motion or stereo.

It turns out that the distortion of perceptual space depends on the direction in the image along which the computation of local correspondence is made. Our analysis leads us to question whether there is any deep reason why the metrics of stereoscopic space and motion space must be different, as some investigators are accustomed to believe [24, 27]. Our unified framework explains the differences between the stereoscopic and motion perceptual spaces on the basis of the image directions along which measurements are made; for stereo, the assumption is that they are made in the horizontal direction, while for motion they are made in any possible direction.

Finally, in the light of the misperceptions arising from stereopsis and motion, the question of how much information we should expect from these modules must be raised. The iso-distortion framework can be used as an avenue for discovering other properties of perceived space. Such properties may lead to new representations of space that can

be examined through further psychophysical studies.

References

[1] J.Y. Aloimonos and D. Shulman. Learning early-vision computations. *Journal of the Optical Society of America A*, 6:908–919, 1989.

[2] L. Cheong, C. Fermüller, and Y. Aloimonos. Interaction between 3D shape and motion: Theory and applications. Technical Report CS-TR-3480, Center for Automation Research, University of Maryland, June 1996.

[3] K. Daniilidis. *On the error sensitivity in the recovery of object descriptions.* PhD thesis, Department of Informatics, University of Karlsruhe, Germany, 1992, in German.

[4] O. Faugeras. *Three Dimensional Computer Vision.* MIT Press, Cambridge, MA, 1992.

[5] C. Fermüller and Y. Aloimonos. Direct perception of three-dimensional motion from patterns of visual motion. *Science*, 270:1973–1976, 22 December 1995.

[6] C. Fermüller and Y. Aloimonos. Ordinal representations of visual space. In *Proc. ARPA Image Understanding Workshop*, pages 897–903, 1996.

[7] C. Fermüller and Y. Aloimonos. Towards a theory of direct perception. In *Proc. ARPA Image Understanding Workshop*, pages 1287–1295, 1996.

[8] J. Foley. Effects of voluntary eye movement and convergence on the binocular appreciation of depth. *Perception and Psychophysics*, 11:423–427, 1967.

[9] J. Foley. Binocular distance perception. *Psychological Review*, 87:411–434, 1980.

[10] J. Gibson. *The Perception of the Visual World.* Houghton Mifflin, Boston, 1950.

[11] W. Hoffman. The Lie algebra of visual perception. *Journal of Mathematical Psychology*, 3:65–98, 1966.

[12] B. Horn. *Robot Vision.* McGraw Hill, New York, 1986.

[13] E.B. Johnston. Systematic distortions of shape from stereopsis. *Vision Research*, 31:1351–1360, 1991.

[14] B. Julesz. *Foundations of Cyclopean Perception.* University of Chicago Press, Chicago, IL, 1971.

[15] J. Koenderink and A. van Doorn. Two-plus-one-dimensional differential geometry. *Pattern Recognition Letters*, 15:439–443, 1994.

[16] J. Koenderink and A. van Doorn. Relief: Pictorial and otherwise. *Image and Vision Computing*, 13:321–334, 1995.

[17] S. Kosslyn. *Image and Brain.* MIT Press, Cambridge, MA, 1993.

[18] R. Luneburg. *Mathematical Analysis of Binocular Vision.* Princeton University Press, Princeton, NJ, 1947.

[19] D. Marr. *Vision.* W.H. Freeman, San Francisco, CA, 1982.

[20] S. Maybank. *Theory of Reconstruction from Image Motion.* Springer, Berlin, 1993.

[21] J. Mayhew and H. Longuet-Higgins. A computational model of binocular depth perception. *Nature*, 297:376–378, 1982.

[22] K. Ogle. *Researches in Binocular Vision.* Hafner, New York, 1964.

[23] J. Semple and L. Roth. *Inroduction to Algebraic Geometry.* Oxford University Press, Oxford, UK, 1949.

[24] J. Tittle, J. Todd, V. Perotti, and J. Norman. Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception and Performance*, 21:663–678, 1995.

[25] J. Todd and P. Bressan. The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception and Psychophysics*, 48:419–430, 1990.

[26] R. Tsai and T. Huang. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:13–27, 1984.

[27] M. Wagner. The metric of visual space. *Perception and Psychophysics*, 38:483–495, 1985.

ORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1996	Technical Report

4. TITLE AND SUBTITLE

Explaining Human Visual Space Distortion

5. FUNDING NUMBERS

N00014-96-1-0587
DAAH04-93-G-0419

6. AUTHOR(S)

Cornelia Fermüller, LoongFah Cheong, and Yiannis Aloimono

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Computer Vision Laboratory
Center for Automation Research
University of Maryland
College Park, MD 20742-3275

8. PERFORMING ORGANIZATION REPORT NUMBER

CAR-TR-833
CS-TR-3662

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research, 800 North Quincy Street, Arlington, VA 22217-5660

Army Research Office, P.O. Box 12211, Research Triangle Park, NC 27709-2211

ARPA, 3701 N. Fairfax Dr., Arlington, VA 22203-1714

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release.
Distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

A number of experiments have recently been conducted to compare aspects of depth judgment due to stereoscopic and monocular motion perception. In these experiments, it has been shown that from stereo vision humans over-estimate depth (relative to fronto-parallel size) at near fixations and under-estimate it at far fixations, whereas human depth estimates from visual motion are not affected by the fixation point. On the other hand, the orientation of an object in space does not affect depth judgment in stereo vision while it has a strong effect in motion vision, for the class of motions tested. This paper develops a computational geometric model that explains why such distortion might take place. The basic idea is that, both in stereo and motion, we perceive the world from multiple views. Given the rigid transformation between the views and the properties of the image correspondence, the depth of the scene can be obtained. Even a slight error in the rigid transformation parameters causes distortion of the computed depth of the scene. The unified framework introduced here describes this distortion in computational terms, in order to explain a number of recent psychophysical experiments on the perception of depth from motion or stereo.

14. SUBJECT TERMS

Visual perception, depth perception, stereopsis, structure from motion, visual space distortion

15. NUMBER OF PAGES

30

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements.***

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	**PR**	- Project
G	- Grant	**TA**	- Task
PE	- Program Element	**WU**	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

- **DOD** - See DoDD 5230.24, "Distribution Statements on Technical Documents."
- **DOE** - See authorities.
- **NASA** - See Handbook NHB 2200.2.
- **NTIS** - Leave blank.

Block 12b. Distribution Code.

- **DOD** - Leave blank.
- **DOE** - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
- **NASA** - Leave blank.
- **NTIS** - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.